PROJECT 4

May The 4s be with You

Wee Jin
Deepankar Sharma
Jerome Sim
Kim Chen

A long time ago, in a galaxy far, far away...

The Jedi High Council, peacekeepers of the galaxy, realised that they'd spent a lot of time and resources resolving conflicts after they had already occurred.

They came to understand that the best way to maintain peace was to pre-empt volatile situations around the galaxy so that they could intervene before any potentially violent outcomes.

“Governments have a history of responding differently toward various types of protests. We will be using classification models to predict the government response - specifically, if it could be aggressive to a given protest.

Being able to accurately predict this will help us prevent any unnecessary economic losses and threats that may threaten the stability of the galaxy.”

—THE JEDI COUNCIL

CONTENTS

1 DATA CLEANING

How we obtain and optimise the data for the rest of the processes

2 EDA & DATA VIS

What are the things that cause a disturbance in The Force?

3 MODELLING

Machines Learning how to sense a disturbance in The Force

4 LIMITS & RECOS

What have we learnt from this, and how can we improve?

1

DATA CLEANING

Understanding the Data

What is it?

Dataset on protests against governments from 162 countries, 1990-2020

How is it collected?

Using newspaper sources from New York Times, Times of London and others

What is the scope?

Protests targeted at state policy only, does not include actions against companies, disputes between groups etc.

Data processing?

Comprehensive feature engineering done to group protester demands, define government responses, compute protest duration and assess text data

15,239

protests on Earth from 1990 - 2020

2 EDA & DATA VIS

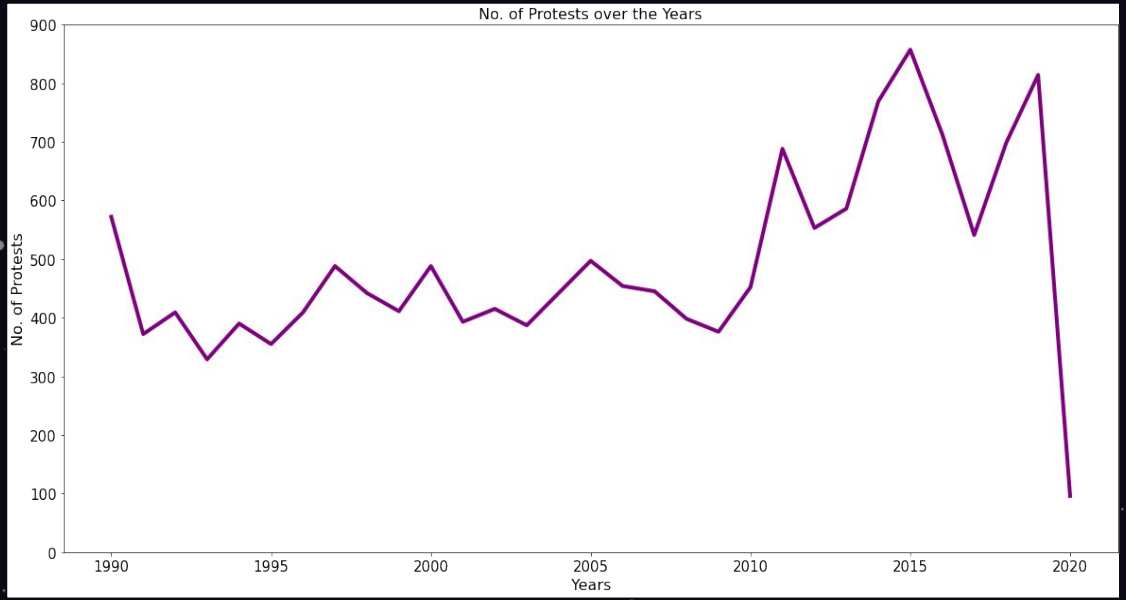
Protest numbers had remained fairly stable till 2010...

- **1990 - 2010**

  Numbers of protests ranged from 300-600 annually.

- **2011 -2020**

  Number of protests surged in 2011, 2015 and 2019. Issues of income inequality and race surfaced.



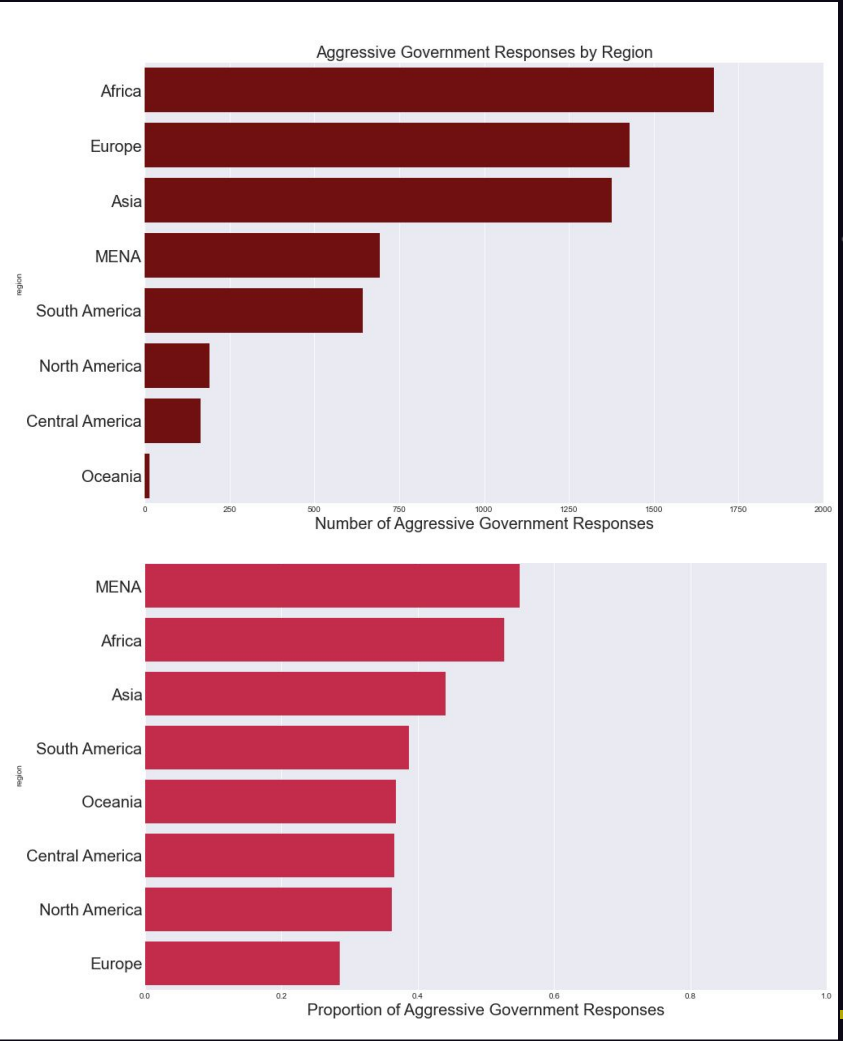
How did Govts from diff regions respond?

- **Aggressive tactics**

  Include killings, shootings, beatings, arrests and crowd dispersals.

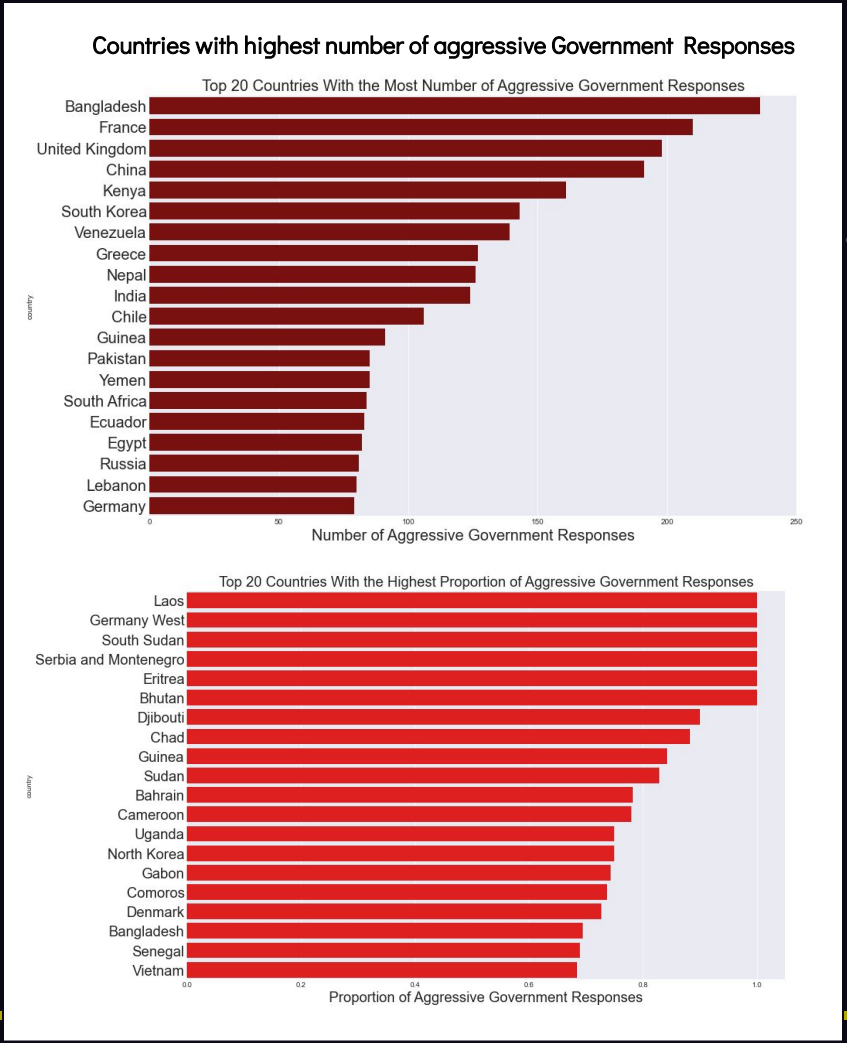
- **Regions with higher aggressive Govt responses**

  Governments from Africa, Asia and the Middle East ranked highest on the aggression scale. Europe most tolerant.



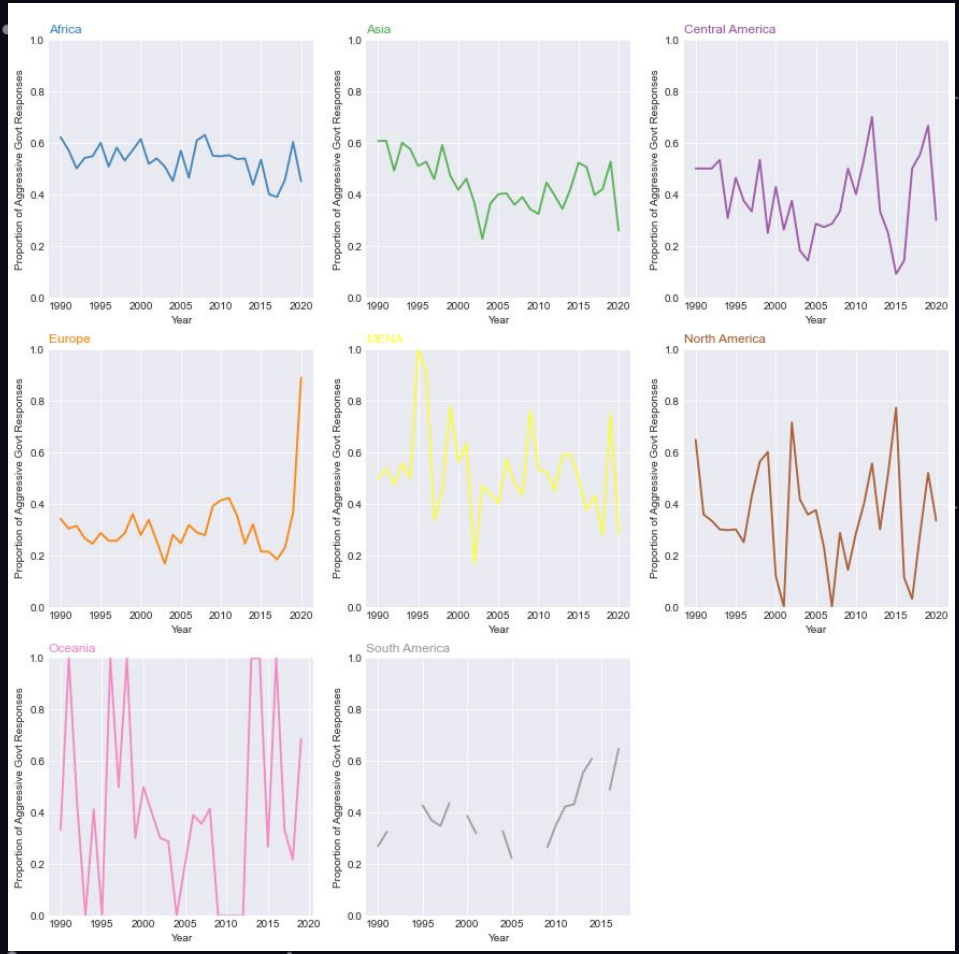
Govts adopting hard-line approach to protests are...

- **Nations with low GDP per Capita**
  More than half of those in the top 20 list for highest proportion of aggressive govt responses.
- **War torn countries**
  Close to one third of countries in the list had armed conflicts over the past 30 years.
- **South Korea? UK? France?**
  Unemployment, religion, Brexit, Climate Change, student protests.



Have things changed over the years?

- **Cent, Nth America, MENA**
  No discernible trends but high fluctuations suggest regime changes.
- **Europe**
  Aggressive Govt reaction to climate crisis and COVID in 2019, 2020.
- **Africa, Asia**
  Governments adopting 'softer' approach in recent years.
- **Oceania, South America**
  Low protest numbers and generally tolerant governments.



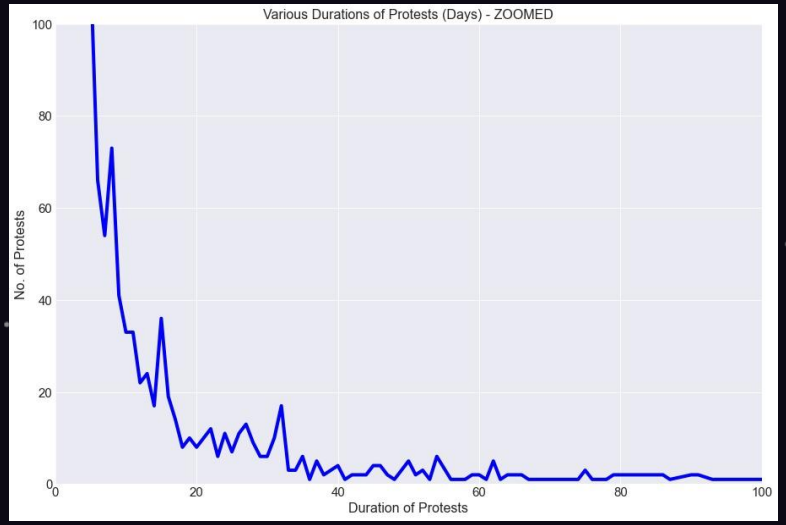
DURATION OF PROTEST

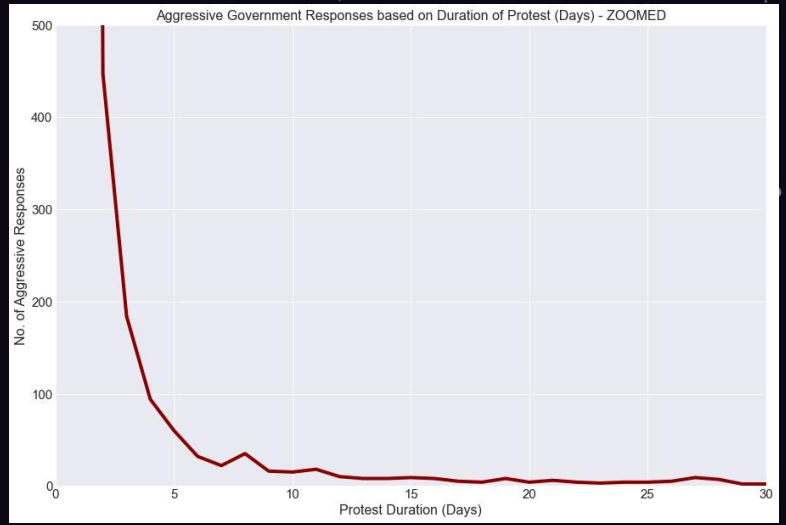
TOTAL PROTESTS

- Most frequent duration: 1 day
- 13,203 out of 15,239 protests (87%)
- Only 729 protests have made it past 100 days

AGGRESSIVE GOVT RESPONSE

- Most frequent duration: 1 day
- 1,414 protests knocked back by end of the day





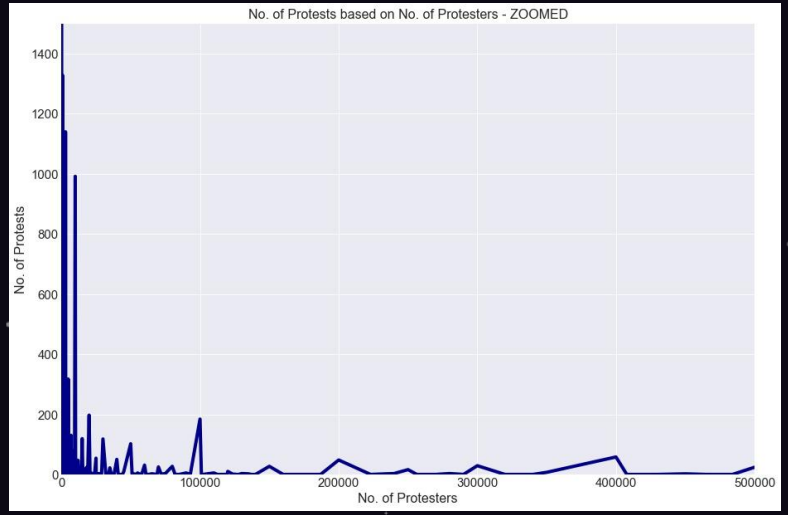
NO. OF PROTESTERS

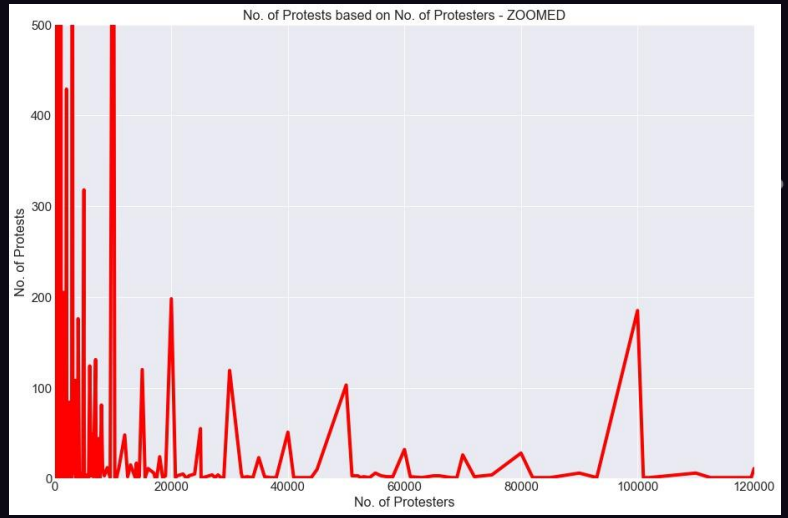
TOTAL PROTESTS

- Most frequent no. of protesters: 50
- 2,987 out of 15,239 protests (20%)
- 100 protesters & 1,000 protesters

AGGRESSIVE GOVT RESPONSE

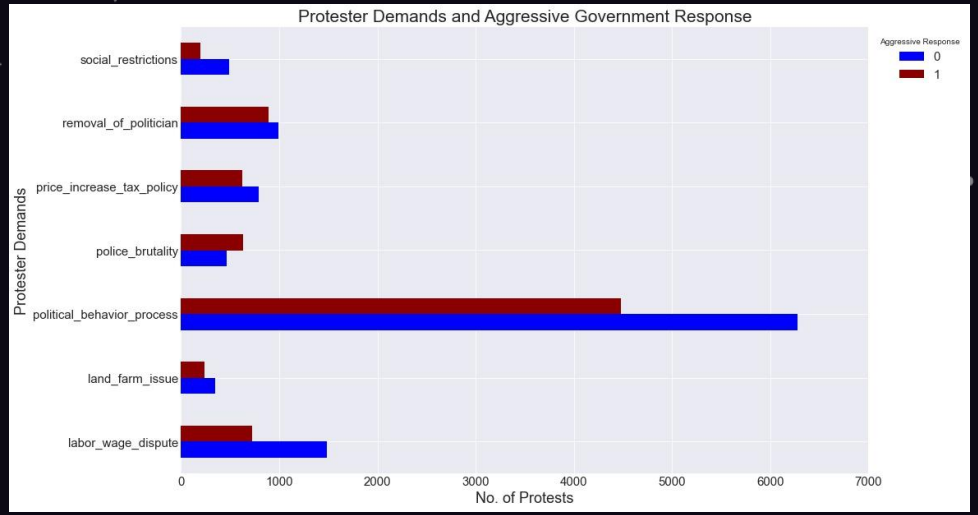
- 1,390 protests featuring 50 protesters (47%)
- 100 protesters: 694 / 1,355 protests (51%)
- 1,000 protesters: 570 / 1,327 protests (43%)





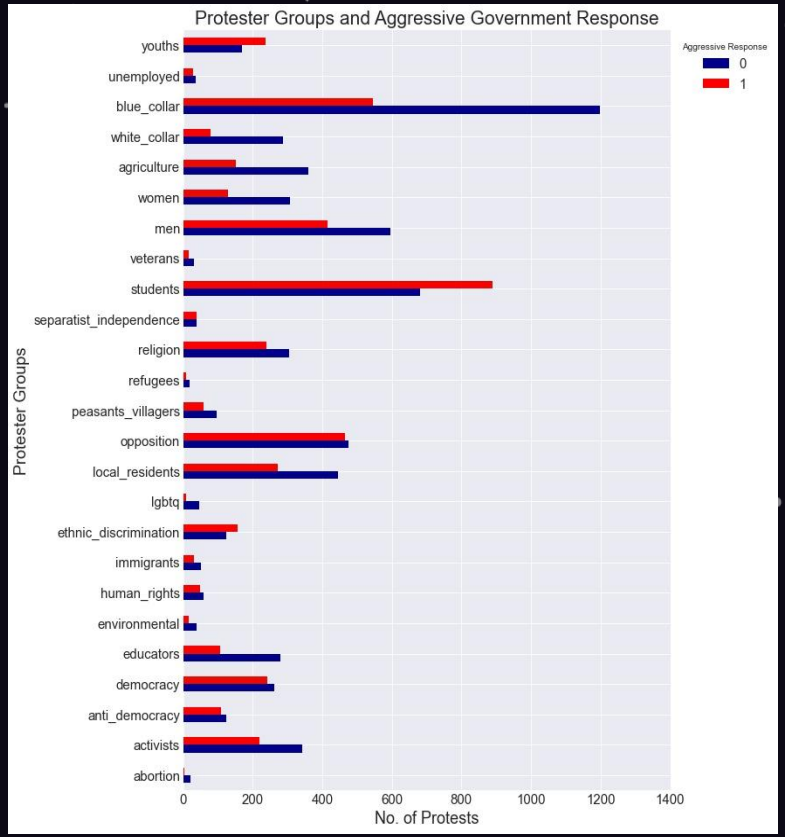
PROTESTER DEMANDS

- **POLITICAL BEHAVIOR**
  - Highest Protests: 10,750
  - Highest Aggressive Responses: 4,475 (42%)
- **REMOVAL OF POLITICIAN**
  - 3rd Highest Protests: 1,889
  - 2nd Highest Aggressive Responses: 894 (47%)
- **POLICE BRUTALITY**
  - 3rd LOWEST Protests: 1,101
  - 634 Aggressive Responses (57%)
- **LABOR WAGE DISPUTE**
  - 2nd Highest Protests: 2,211
  - 3rd Highest Aggressive Responses: 723 (32%)



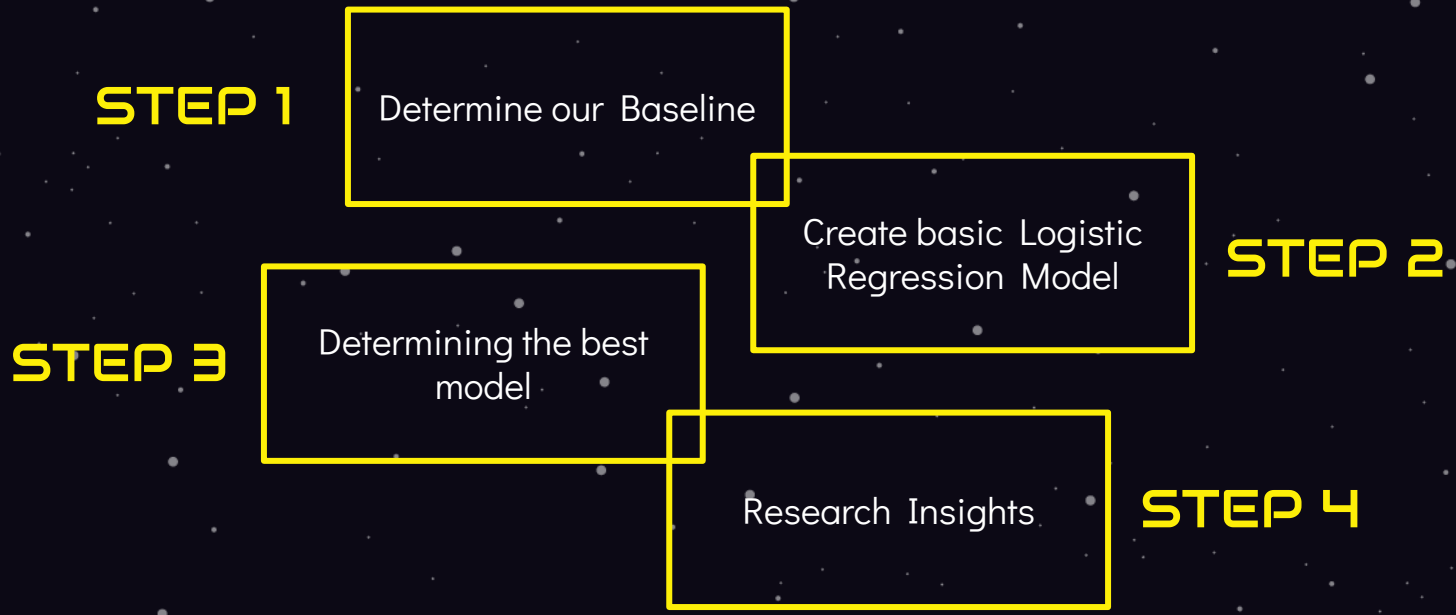
PROTESTER GROUPS

- **BLUE COLLAR**
  - Highest Protests: 1,744
  - Aggressive Responses: 546 (31%)
- **STUDENTS**
  - 2nd Highest Protests: 1,572
  - Highest Aggressive Responses: 890 (57%)
- **OPPOSITION**
  - Protests: 940
  - 3rd Highest Aggressive Responses: 464 (49%)
- **DEMOCRACY**
  - 48% Aggressive Responses
- **YOUTHS**
  - HIGHEST Aggressive Responses (72%)



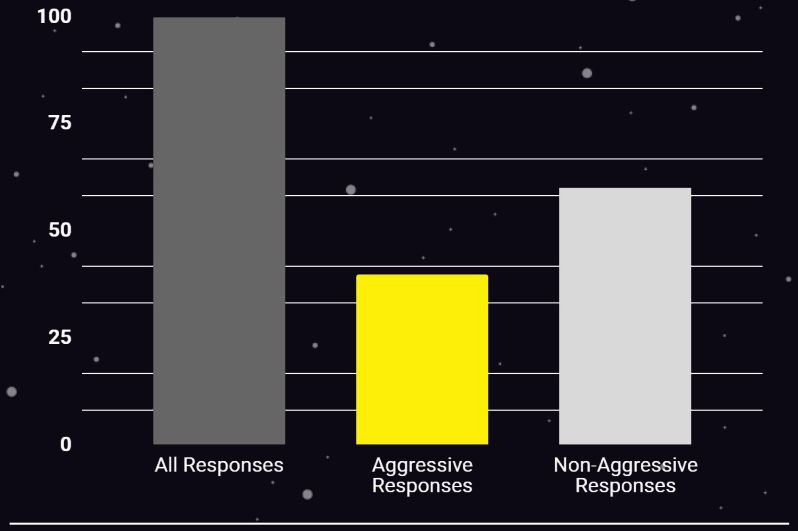
MODELLING

MODELLING PROCESS



OUR BASELINE

- **All Responses**
  100% of the responses to protests
- **Aggressive Responses**
  ~40% of responses are aggressive
- **Non-Aggressive Responses**
  ~60% of responses are non-aggressive



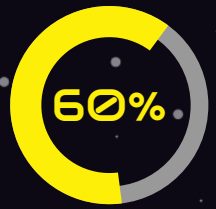
BASIC LOGISTIC REGRESSION MODEL

TRAIN 79%

TEST 79%

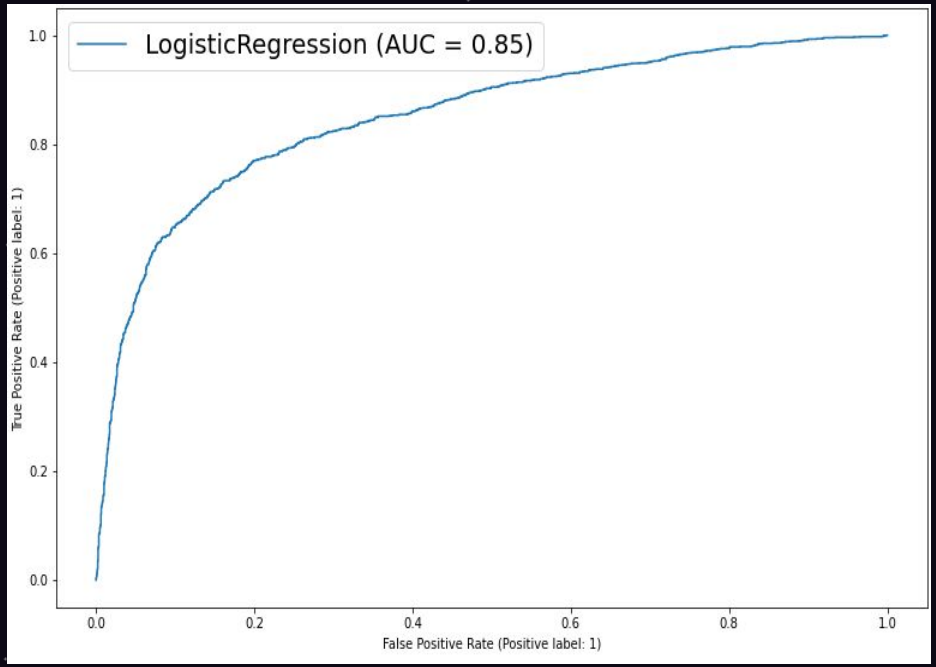
Specificity

Ability to classify non-aggressive responses



Sensitivity

Ability to classify aggressive responses



DETERMINING THE BEST MODEL

	Train Score	Test Score	SSB
LR	0.794	0.797	0.737
KNN	0.785	0.753	0.683
Decision Tree	0.782	0.790	0.643
Bagging(DT)	0.917	0.801	0.765
Bagging(LR)	0.906	0.801	0.765

RESEARCH INSIGHTS

Protestorviolence

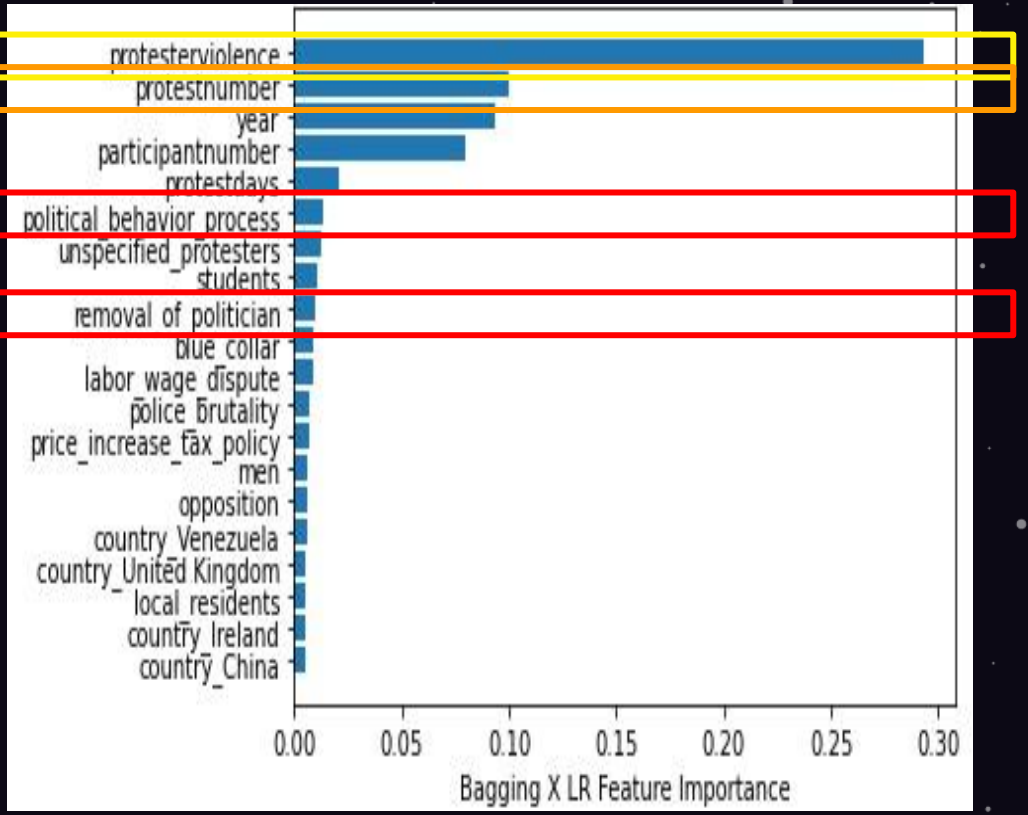
Increases probability of an aggressive response the most

Protestnumber

Governments' tolerance decreases with the number of protests in a year

Political Demands

Aggression more likely when protests are motivated by political change



4 LIMITATIONS & RECOMMENDATIONS

LIMITATIONS

1. The existing dataset draws heavily from western media. Gathering data from an array of international sources would reduce the inherent bias from the current data collection process.

2. The ‘notes’ section in the data was not specific enough and did not provide much insight. Hence more datasets can be included in order to help supplement this area.

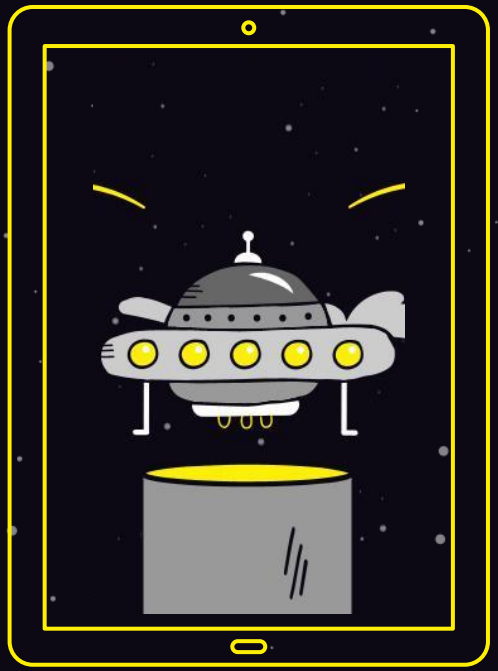
3. Our model had an unchanging inability of every optimally-tuned model to classify aggressive responses as well as non-aggressive responses.

RECOMMENDATIONS

1. Across the galaxy our observers should immediately flag and escalate to the Jedi Council any protests that see its protesters employing violence.

2. Include other forms of data sets
    a. Social Media data

RECOMMENDATIONS (Future Projects)



Protest Database System (PDS)

This database will function as an automated system with a goal to replace the labor-intensive process of having human coders look for information about protests in news sources with a computer-aided set of protocols.

Video Analysis with OpenCV

Object Detection, Object Recognition, Object Tracking

THANKS!

DO YOU HAVE ANY QUESTIONS?